

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号
特開2002-140332
(P2002-140332A)

(43) 公開日 平成14年 5月17日 (2002.5.17)

(51) IntCl. ⁷	識別記号	F I	キーワード (参考)
G 0 6 F 17/30	1 7 0 3 5 0	G 0 6 F 17/30	1 7 0 B 5 B 0 5 0 3 5 0 C 5 B 0 7 5
G 0 6 T 1/00 7/00	2 0 0 3 0 0	G 0 6 T 1/00 7/00	2 0 0 A 5 L 0 9 6 3 0 0 F

審査請求 未請求 請求項の数10 O L (全 11 頁)

(21) 出願番号 特願2000-336374(P2000-336374)

(22) 出願日 平成12年11月 2 日 (2000. 11. 2)

(71) 出願人 000004226

日本電信電話株式会社

東京都千代田区大手町二丁目 3 番 1 号

(72) 発明者 谷口 展郎

東京都千代田区大手町二丁目 3 番 1 号 日
本電信電話株式会社内

(74) 代理人 100073760

弁理士 鈴木 誠

F ターム (参考) 5B050 EA18 GA08

5B075 ND06 NK02 NK06 PP22 PQ02

PRO6 PRO8 QM08

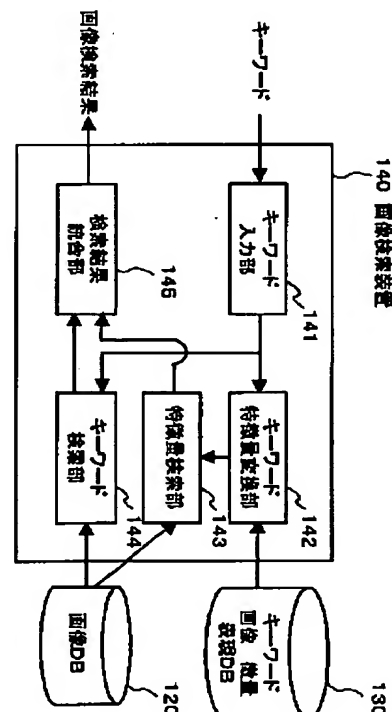
5L096 BA08 JA11 KA09

(54) 【発明の名称】 特徴量重要度算出方法、それを用いたキーワード画像特徴量表現データベース作成及び画像データベース検索

(57) 【要約】

【課題】 適切なキーワードが付与されていないような画像データベースでも高精度の検索を可能にする。

【解決手段】 画像ごとに、キーワードと特徴量セットを関連づけて蓄積した画像データベース120と、キーワードごとの画像特徴量表現（特徴量セットと各特徴量の重要度）を蓄積したキーワード画像特徴量表現データベース130を備える。画像検索装置140は、入力されたキーワードによりキーワード画像特徴量表現データベース130を検索して当該キーワードの画像特徴量表現を取得し、該画像特徴量表現により画像データベース120を類似検索して類似度の大きい画像を取得し、また、入力されたキーワードにより画像データベース120を検索して、当該キーワードに対応する画像を取得し、これら取得した画像から最終的な検索結果画像を決定して出力する。



【特許請求の範囲】

【請求項 1】 画像間の類似度の算出で用いられる特徴量（特徴ベクトル）の重要度を算出する特徴量重要度算出方法であって、

対象とする特徴量ベクトル群の各次元における最大値と最小値を求めて、該最大値と最小値が等しい次元を除去し、

前記除去した次元の残りの次元（次元数 n とする）からなる特徴量ベクトル群における外接長矩形の面積を求め、

前記面積の n 乗根の逆数を特徴量の重要度とすることを特徴とする特徴量重要度算出方法。

【請求項 2】 コンピュータ読み取り可能な記録媒体であって、請求項 1 記載の特徴量重要度算出方法をコンピュータで実行するためのプログラムを記録したことを特徴とする記録媒体。

【請求項 3】 画像ごとに、1 以上のキーワードと特徴量セットが関連づけられて蓄積された画像データベースを用いて、キーワードごとの画像特徴量表現（特徴量セットと各特徴量の重要度）を作成するキーワード画像特徴量表現作成方法であって、

前記画像データベースから、キーワードを抽出して、当該キーワードが付与された画像の特徴量セット群を取得し、

前記特徴量セット群について、各特徴量ごとに平均値を算出して、当該キーワードの画像特徴量表現の特徴量セットとし、

前記特徴量セット群について、各特徴量ごとに請求項 1 記載の特徴量重要度算出方法を適用して重要度を算出し、当該キーワードの画像特徴量表現の各特徴量の重要度とすることを特徴とするキーワード画像特徴量表現作成方法。

【請求項 4】 請求項 3 記載のキーワード画像特徴量表現作成方法において、

画像データベースに、画像の部分画像ごとにキーワードが付与されている場合、キーワードの画像特徴量表現に加えて、キーワードの部分画像特徴量表現も作成することを特徴とするキーワード画像特徴量表現作成方法。

【請求項 5】 画像ごとに、1 以上のキーワードと部分画像が関連づけられて蓄積された画像データベースについて、部分画像にキーワードを追加して画像データベースを再構築する画像データベース構築方法であって、追加すべきキーワードにより画像データベースを検索し、

検索結果の画像の所望部分を指定することで、その近辺に位置する部分画像を取得し、

取得された部分画像の中から所望の部分画像を指定することで、当該部分画像にキーワードを追加することを特徴とする画像データベース構築方法。

【請求項 6】 画像ごとに、1 以上のキーワードと特徴

量セット、部分画像と特徴量セットが関連づけられて蓄積された画像データベースについて、部分画像にキーワードを追加して画像データベースを再構築する画像データベース構築方法であって、

あらかじめキーワードごとの部分画像特徴量表現（特徴量セットとその各特徴量の重要度）が蓄積されたキーワード画像特徴量表現データベースを用い、

追加すべきキーワードによりキーワード画像特徴量表現データベースを検索して、当該キーワードの部分画像特徴量表現を取得し、

前記取得した部分画像特徴量表現により、画像データベースの各部分画像の特徴量セットとの類似度検索を行って類似度の高い部分画像を取得し、

前記取得された部分画像の中から所望の部分画像を指定することで、当該部分画像にキーワードを追加することを特徴とする画像データベース構築方法。

【請求項 7】 画像ごとに、1 以上のキーワードと特徴量セットが関連づけられて蓄積された画像データベースから、キーワードを入力して画像を検索する画像検索方法であって、

キーワードごとの画像特徴量表現（特徴量セットとその各特徴量の重要度）が蓄積されたキーワード画像特徴量表現データベースを用い、

入力されたキーワードによりキーワード画像特徴量表現データベースを検索して、当該キーワードの画像特徴量表現を取得し、

前記取得した画像特徴量表現の特徴量セットと画像データベースの各画像の特徴量セットとの類似度を、画像特徴量表現の重要度を用いて算出し、類似度の大きい画像（以下、第 1 の画像という）を取得し、

入力されたキーワードにより画像データベースを検索して、当該キーワードに対応する画像（以下、第 2 の画像という）を取得し、

前記取得した第 1 の画像と第 2 の画像から最終的な検索結果画像を決定して出力することを特徴とする画像検索方法。

【請求項 8】 コンピュータ読み取り可能な記録媒体であって、請求項 7 記載の画像検索方法をコンピュータで実行するためのプログラムを記録したことを特徴とする記録媒体。

【請求項 9】 キーワードごとの画像特徴量表現（特徴量セットの各特徴量の重要度）を蓄積したキーワード画像特徴量表現データベースを作成するキーワード画像特徴量表現データベース作成システムであって、

画像ごとに、1 以上のキーワードと特徴量セットが関連づけられて蓄積された画像データベースと、

前記画像データベースから、キーワードを抽出して、当該キーワードが付与された画像の特徴量セット群を取得する手段と、

前記特徴量セット群から、各特徴量ごとに平均値を求め

て特徴量セットを算出する手段と、
前記特徴量セット群から、各特徴量の重要度を算出する手段と、

前記算出された特徴量セットと各特徴量の重要度の組をキーワードの画像特徴量表現としてキーワード画像特徴量表現データベースに格納する手段と、を有することを特徴とするキーワード画像特徴量表現データベース作成システム。

【請求項10】 画像ごとに、1以上のキーワードと特徴量セットが関連づけられて蓄積された画像データベースと、

キーワードごとの画像特徴量表現（特徴量セットと各特徴量の重要度）を蓄積したキーワード画像特徴量表現データベースと、

検索したい画像を表わすキーワードを入力する手段と、
入力されたキーワードにより前記キーワード画像特徴量表現データベースを検索して、当該キーワードの画像特徴量表現を取得する手段と、

前記取得した画像特徴量表現の特徴量セットと前記画像データベースの各画像の特徴量セットとの類似度を、画像特徴量表現の重要度を用いて算出し、類似度の大きい画像（第1の画像）を取得する手段と、

入力されたキーワードにより前記画像データベースを検索して、当該キーワードに対応する画像（第2の画像）を取得する手段と、

前記取得した第1の画像と第2の画像から最終的な検索結果画像を決定して出力する手段と、
を有することを特徴とする画像検索システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明、画像処理システム、画像検索システム等において画像間の類似度の算出に用いられる画像特徴量（特徴ベクトル）の重み付け技術、それを用いたキーワード画像特徴量表現データベース作成技術および画像データベース検索技術等に関する。

【0002】

【従来の技術】画像検索の従来技術は、検索したい画像のサンプル画像を入力し、それと類似した画像を画像データベースから検索する方式及び検索したい画像の特徴を表すキーワードを入力し、それを表す画像を画像データベースから検索する方式に大別される。また、類似検索で画像間の類似度の算出に用いられる画像特徴量の重要度（重み係数）の算出法の従来技術は、分散あるいは標準偏差の逆数を利用する方法及び、面積の逆数を利用する方法に大別される。以下、これらについて概要を説明する。

【0003】図1は従来のサンプル画像による画像検索システムの概略構成を示す。図1において、画像検索装置10は特徴量抽出部11、類似度算出部12、画像出力部13からなる。画像データベース20には、図2に

示すように、各画像とその特徴量セットが関連づけられて蓄積されている。特徴量セットは複数の特徴量（色、形、模様など）の組からなり、各特徴量は複数次元の特徴ベクトルで表わされる。特徴量抽出部11は、入力されたサンプル画像から特徴量（特徴量セット）を抽出する。類似度算出部12は、抽出されたサンプル画像の特徴量セットと画像データベース20に格納された各画像の特徴量セットとの類似度を、予め定めた特徴量毎の重要度（重み係数）を加味して計算し、類似度が大きい順に一定数あるいは一定割合の画像を画像データベース20から読み出す。画像出力部13は、この読み出された画像データを検索結果として出力する。

【0004】図3は従来のキーワードによる画像検索システムの概略構成を示す。図3において、画像検索装置30はキーワード検索部31、画像出力部32からなる。画像データベース40には、図4に示すように、各画像とその特徴を表わす1以上のキーワードが関連づけられて蓄積されている。利用者が検索したい画像の特徴を表すキーワード（リンゴ、花瓶など）を入力すると、キーワード検索部31は、入力されたキーワードと画像データベース40に格納された各画像のキーワードを照合し、入力キーワードと一致するキーワードを持つ画像を画像データベース40から読み出す。画像出力部32は、この読み出された画像データを検索結果として出力する。

【0005】次に、類似検索で画像間の類似度の算出に用いられる画像特徴量の重要度の従来の算出方法について説明する。

【0006】一方の画像*i*の特徴量（*A*, *B*, ...）の組を $V_i = \{v_{iA}, v_{iB}, \dots\}$ とし、他方の画像*j*の特徴量（*A*, *B*, ...）の組を $V_j = \{v_{jA}, v_{jB}, \dots\}$ とすると、 V_i と V_j の類似度 $S(V_i, V_j)$ は、 $S(V_i, V_j) = w_A \times d(v_{iA}, v_{jA}) + w_B \times d(v_{iB}, v_{jB}) + \dots$ として求まる。ここで、 w_A, w_B, \dots は特徴量毎の重要度（重み係数）、 $d(*, *)$ は特徴量（特徴ベクトル）毎の類似度である。

【0007】特徴量毎の重要度（重み係数）を求める従来技術の第1の方法は、特徴量毎のベクトル群（特徴量ベクトル群）の分散（あるいは標準偏差）の逆数を利用するものである。一般にベクトル群 $V = \{v_1, \dots, v_n\}$ が与えられたとき、その分散*s*および標準偏差*d*は、 V の重心ベクトルを V_g として、

$$s = (|v_1 v_g|^2 + \dots + |v_n v_g|^2) / n$$

$$d = \sqrt{s}$$

で求まる。*s*あるいは*d*が小さければ小さいほど、ベクトル群*V*内の各ベクトルは空間中のある特定の領域に集中して存在し、逆に大きければ大きいほどばらばらに散らばって存在しているといえる。図5は、この様子を示したもので、丸印がベクトル群内の各ベクトル、バツ印

がベクトル群の重心である。

【0008】いま、あるキーワードWに対応する特徴量毎のベクトル群が、特徴量Aにおいて $V_A = \{v_{A1}, v_{A2}, \dots\}$ 、特徴量Bにおいて $V_B = \{v_{B1}, v_{B2}, \dots\}$ として与えられるものとする。

【0009】与えられるベクトル群の分散（あるいは標準偏差）が小さい場合は、その特徴量は、キーワードと他のキーワードを識別するうえで重要であるということを意味する（図5（A）のケース）。逆に分散（あるいは標準偏差）が大きい場合は、その特徴量は、そのキーワードと他のキーワードを識別するうえであまり重要でないといえる（図5（B）のケース）。これに着目して、例えば、特徴量Aのベクトル群 V_A の分散（あるいは標準偏差が s_A （あるいは d_A ）、特徴量Bのベクトル群 V_B の分散（あるいは標準偏差）が s_B （あるいは d_B ）であるとき、検索時に各特徴量A、Bに与える重要度（重み係数） w_A 、 w_B をそれぞれ $1/s_A$ （あるいは $1/d_A$ ）、 $1/s_B$ （あるいは $1/d_B$ ）とするものである。他の特徴量についても同様である。

【0010】特徴量毎の重要度（重み係数）を求める従来技術の第2の方法は、特徴量毎のベクトル群の面積の逆数を利用するものである。同様に、ベクトル群 $V = \{v_1, \dots, v_n\}$ が与えられたとき、その外接多次元矩形Rあるいは外接多次元円Sなどの面積をaとすると、一般に、aが小さければ小さいほど、ベクトル群V内の各ベクトルは空間中のある特定の領域に集中して存在し、逆に大きければ大きいほどばらばらに散らばって存在しているといえる。図6及び図7は、それぞれ外接多次元矩形と外接多次元円について、この様子を示したものである。

【0011】これに着目して、例えば、特徴量Aのベクトル群 V_A の面積が a_A 、特徴量Bのベクトル群 V_B の面積が a_B であるとき、検索時に各特徴量A、Bに与える重要度（重み係数） w_A 、 w_B を、それぞれ $1/a_A$ 、 $1/a_B$ とするものである。他の特徴量についても同様である。

【0012】

【発明が解決しようとする課題】従来技術において、サンプル画像を用いる画像検索は、利用者にとって、サンプルとなる画像を入力したり指定するのがかなり煩わしい問題がある。

【0013】また、この類似検索時、入力されたサンプル画像と画像データベースに格納された画像間の類似度の算出に用いられる特徴量毎の重要度（重み係数）として、分散や標準偏差を利用した場合、一般に分散や標準偏差は、ベクトルが存在する空間の次元数が大きくなるほど大きな値をとる傾向があるため、異なる次元数の特徴量空間における分散（あるいは標準偏差）を単純に同列に扱うことはできず、次元数の違いを無視して用いると、低次元の特徴量を重視しすぎることになり、検索度

の低下を導びく問題がある。同様に、面積を利用した場合も、次元数の異なる空間における値を単純に同列に扱うことはできない。更に、外接矩形面積については、ある特定の次元軸の辺長がゼロになった場合、たとえ他の次元軸における散らばりが大きくても、面積自体は必ずゼロになるという問題がある。

【0014】従来技術において、キーワードを用いる画像検索は、必ずしも画像データベース内の全ての画像に、必要なキーワードのすべてが付与されているとは限らず、また、キーワードを付与する人間によって、キーワードにばらつきが生じ、高精度の検索を期待できない問題がある。

【0015】本発明は、上述したような従来技術の問題点を解決した画像特徴量の重要度算出方法、それを用いたキーワード画像特徴量表現データベース作成、画像データベース構築及び画像検索システム等を提供することを目的とする。

【0016】

【課題を解決するための手段】本発明の特徴量の重要度算出方法は、特徴量毎にベクトル群が与えられたとき、その外接多次元矩形面積を基本とするが、辺長がゼロになる次元は除外し、全ての辺長がゼロになる場合を除いて、必ず外接矩形面積 s が $s > 0$ となるようにする。そして、この辺長がゼロでない次元の数を n として、 $w = 1/s^{1/n}$ つまり、外接多次元矩形面積 s の n 乗根の逆数を、特徴量毎の重要度（重み係数）とする。これは、辺長がゼロでない次元のみで構成された空間（これを有効部分空間と称す）上の多次元矩形を多次元正方形と見做したときの一辺の長さに相当する。図8は、この様子を示したものである。

【0017】本発明の特徴量毎の重要度算出方法では、辺長がゼロでない次元のみで構成された有効部分空間の次元数で冪開しているため、元々の特徴量空間の次元数の行程の影響を受けにくい重要度（重み係数）の生成が可能である。

【0018】次に、本発明は、各画像に対して1以上のキーワードと特徴量セットが付与された画像データベースから、上記特徴量の重要度算出方法を用いて、キーワード毎の画像特徴量表現（特徴量セットとその各特徴量の重要度）を算出して、キーワードを特徴量に変換するためのキーワード画像特徴量表現データベースを自動構築することを特徴とする。また、本発明は、このようなキーワード画像特徴量表現データベース等を用いて部分画像にキーワードを付与し、画像データベースを再構築することを特徴とする。

【0019】さらに、本発明は、キーワード画像特徴量表現データベースと画像データベースを組み合わせ、入力されたキーワードをキーワード画像特徴量表現データベースを用いて特徴量表現に変換して、キーワードと特徴量表現の両者を用いて画像データベースを検索する画像

検索システムを構築することを特徴とする。これにより、サンプル画像が不要であり、また、適切なキーワードが付与されない画像データベースでも、高精度の画像検索が可能である。

【0020】

【発明の実施の形態】以下、本発明の実施の形態について図面により詳しく説明する。

〈特徴量の重要度算出〉図9は、本発明による特徴量の重要度算出手順を示す。一例として、元々の特徴量空間の次元数を3（3次元：x, y, z）とし、特徴量Aにおけるベクトル群として、特徴量A1 = (x_{A1}, y_{A1}, z_{A1})、特徴量A2 = (x_{A2}, y_{A2}, z_{A2})、特徴量A3 = (x_{A3}, y_{A3}, z_{A3}) が与えられたとする。

【0021】まず、特徴量空間の各次元における座標値の最大値、最小値を求める（ステップ101）。例では、x_{A1}, x_{A2}, x_{A3}中の最大値x_{max}と最小値x_{min}, y_{A1}, y_{A2}, y_{A3}中の最大値y_{max}と最小値y_{min}, z_{A1}, z_{A2}, z_{A3}中の最大値z_{max}と最小値z_{min}を求める。

【0022】次に、座標値の最大値と最小値が等しい次元を求める（ステップ102）。すなわち、辺長がゼロになる次元を求める。ここでは、y_{max} = y_{min}とし、y次元が辺長ゼロであったとする。次に、座標値の最大値 = 最小値の次元、すなわち、辺長がゼロとなる次元を除去し、残った特徴量空間（有効部分空間）の次元数nを求める（ステップ103）。例では、n = 2である。

【0023】次に、有効部分空間における特徴量分布の外接多次元長矩形の面積sを算出する（ステップ104）。例では、s = (x_{max} - x_{min}) · (z_{max} - z_{min}) として求まる。従来技術では、y_{max} = y_{min}の場合、s = (x_{max} - x_{min}) · (y_{max} - y_{min}) · (z_{max} - z_{min}) = 0となり、たとえx_{max} ≠ x_{min}, z_{max} ≠ z_{min}であっても、面積sは必ずゼロになるという問題が生じる。

【0024】最後に、面積sのn乗根の逆数を求め、各特徴量の重要度（重み係数）wとする（ステップ105）。例では、w_A = 1 / s^(1/2)を特徴量A1, A2, A3に対する重要度（重み係数）とする。

【0025】ここでは、分かり易いように、特徴量（ベクトル）を3次元とし、対応する特徴量もA1, A2, A3の3個のみとしたが、実際には特徴量は多次元であり、対応する特徴量はそれぞれかなりの数になることは云うまでもない。

【0026】〈キーワード画像特徴量表現データベースの作成〉図10は、本発明によるキーワード画像特徴量表現データベース作成システムの構成例を示す。図10において、キーワード画像特徴量表現データベース作成装置110はキーワード抽出部111、特徴量セット抽出部112、特徴量セット算出部113、重要度算出部114、データ格納部115からなる。画像データベ-

ス120には、図11に示すように、各画像と、その特徴を表わす1以上のキーワード及び特徴量セットが関連づけられて蓄積されている。キーワード特徴量表現データベース130には、図12の下段に示すように、キーワード毎に作成された画像特徴量表現データが蓄積される。特徴量表現データは特徴量セットとその各特徴量の重要度（重み係数）の組で構成される。

【0027】キーワード抽出部111は、画像データベース120から全てのキーワードを抽出する。特徴量セット抽出部112は、キーワード抽出部111で抽出された各キーワードにより画像データベース120を検索して、各キーワード毎に、当該キーワードが付与された全ての画像の特徴量セットを抽出する。図12の上段は、キーワード1に対して画像1、画像2、画像3、…の特徴量セット1、特徴量セット2、特徴量セット3、…が抽出されることを示している。

【0028】特徴量セット算出部113は、各キーワード毎に、当該キーワードの画像特徴量表現における特徴量セットを算出する。具体的には、特徴量セット抽出部112で抽出された当該キーワードに対する各画像の特徴量セットから、対応する特徴量毎、各次元対応に平均値を求める。例えば、図12の下段のキーワード1の画像特徴量表現において、特徴量セット中の特徴量Aは、図12の上段の当該キーワード1に対する画像1、画像2、画像3、…の特徴量セット1、特徴量セット2、特徴量セット3、…における対応する特徴量A1, A2, A3, …について、各次元対応に平均値を算出して得られる。特徴量B, C, Dについても同様である。

【0029】重要度算出部114は、各キーワード毎に、当該キーワードの画像量特徴量表現における各特徴量の重要度を算出する。具体的には、特徴量セット抽出部112で抽出された当該キーワードに対する各画像の特徴量セットから、本発明による特徴量重要度算出方法を用いて各特徴量の重要度（重み係数）を算出する。例えば、図12の下段のキーワード1の画像特徴量表現において、特徴量Aの重要度w_Aは、特徴量A1, A2, A3, …を対象ベクトル群として、図9の手順で算出する。特徴量B, C, Dの各重要度w_B, w_C, w_Dについても同様である。

【0030】なお、キーワードの画像特徴量表現の各特徴量の重要度は、図9で説明した有効部分空間における外接長矩形の面積の代わりに、従来技術の分散や標準偏差あるいは外接矩形面積や外接円面積を用いて算出してもよい。

【0031】データ格納部115は、特徴量セット算出部113および重要度算出部114の算出結果を組み合わせ、順次、キーワードごとの画像特徴量表現（特徴量セットと各特徴量の重要度）をキーワード特徴量表現データベース130に格納する。この結果、キーワード画像特徴量表現データベース130には、図12の下段に

示すようなキーワードの画像特徴量表現が、当該キーワードと関連づけられて蓄積される。

【0032】〈画像検索システム〉図13は、本発明による画像検索システムの構成例を示す。本画像検索システムは画像データベース120、キーワード画像特徴量表現データベース130、画像検索装置140などを備える。画像データベース120は、図11に示したように、各画像と、その特徴量を表わす1以上のキーワードと特徴量セットを関連づけて蓄積している。キーワード画像特徴量表現データベース130は、図12の下段に示したように、キーワードごとの画像特徴量表現（特徴量セットと各特徴量の重要度）を、当該キーワードと関連づけて蓄積している。画像検索装置140はキーワード入力部141、キーワード特徴量変換部142、特徴量検索部143、キーワード検索部144、検索結果統合部145からなる。

【0033】利用者は、検索したい画像の特徴を表わすキーワードを入力する。キーワード入力部141は、入力されたキーワードをキーワード特徴量変換部142とキーワード検索部144に送る。キーワード特徴量変換部142は、入力されたキーワードを用いてキーワード画像特徴量表現データベース130を検索し、該入力されたキーワードに対応する画像特徴量表現（特徴量セットと各特徴量の重要度）を取得する。特徴量検索部143は、キーワード特徴量変換部142で取得した画像特徴量表現により画像データベース120を検索する。具体的には、特徴量検索部143では、画像データベース120の各画像の特徴量セットと画像特徴量表現の特徴量セットとの類似度を、画像特徴量表現の各特徴量の重要度（重み係数）を用いて算出し、類似度が大きい順に一定数あるいは一定割合の画像データを画像データベース120から読み出す。キーワード検索部144は、キーワード入力部141からの入力キーワードにより画像データベース120を検索し、入力キーワードと一致するキーワードを持つ1つあるいは複数の画像データを読み出す。検索結果統合出力部145は、特徴量検索部143による検索結果とキーワード検索部144による検索結果から最終的な検索結果を決定し、その画像データを出力する。

【0034】ここで、検索結果統合出力部145から出力される画像データとしては、（1）特徴量検索部とキーワード検索部での検索結果のすべて（OR条件）、

（2）特徴量検索部とキーワード検索部の両方での検索結果（AND条件）、（3）特徴量検索部での検索結果のみ、（4）キーワード検索部での検索結果のみ、

（5）（1）～（4）について予め定めた順に出力する、などが考えられる。そして、これらのいずれを利用するかは、利用者が任意に選択できるようにする。

【0035】〈画像データベースの構築〉一つの画像内には複数の物体が写っているなど、複数の部分画像に分

けることができる。例えば、林檎と梨と花瓶が置かれたテーブルの昼の情景の画像に、「林檎」、「梨」、「テーブル」、「静物」、「昼」というキーワードがつけられている場合、具象物である「林檎」、「梨」、「テーブル」は、それぞれ画像の特定の部分（部分画像）を表わしている。したがって、部分画像とキーワードを結びつけることができれば、より柔軟かつ精度の高い画像データベースやキーワード画像特徴量表現データベース、およびそれを用いた画像検索システムを構築できる。

【0036】図14は、このような画像データベース構築の概念図を示したものである。図14の左側は、画像データベースの既存のデータ構成であり、画像に対して、1以上のキーワードと特徴量セットが付与され、また、部分画像に対してはその特徴量セットが付与されている。ここでは、この部分画像に対してキーワードを追加し、図14の右側のようなデータ構成とするものである。なお、部分画像には、すでに一部のキーワードが付与されていてもよい。

【0037】図14の右側に示すデータ構成の画像データベースを適用することにより、図15に示すように、より適切なキーワードごとの画像特徴量表現（部分画像特徴量表現）を作成することができる。また、このキーワードの部分画像特徴量表現は、後述するように、さらに部分画像にキーワードを追加する際にも利用できる。

【0038】図16に、部分画像にキーワードを追加する方法の一例を示す。これは、図14に示したように、画像データベースの既存のデータを直接利用するものである。まず、利用者が、追加しようとするキーワードを入力することで、画像データベースから該キーワードの付与された画像を検索する（ステップ201）。検索結果の画像は表示装置に表示される。次に、利用者は、この検索結果の画像から、当該キーワードに関係すると思われる部分を指定し（ステップ202）、画像データベースから、その部分近辺に位置する部分画像を取得する（ステップ203）。次に、利用者は、この取得された部分画像の中から該キーワードに関係すると思われるものを指定する（ステップ204）。この結果、画像データベースの当該画像に対して、指定された部分画像のキーワードリストに該キーワードが追加される（ステップ205）。なお、追加するキーワードは、最初の検索に利用したものと同一でなくてもよく、利用者が追加すべきキーワードを別に入力することでもよい。

【0039】図17に、部分画像にキーワードを追加する方法の別の一例を示す。これは、図15に示すような、すでに一部作成済みのキーワードの部分画像特徴量表現を利用するものである。

【0040】まず、利用者が、追加しようとするキーワードを入力することで、キーワード画像特徴量表現データベースから該キーワードの部分画像特徴量表現（特徴量セットと各特徴量の重要度）を取得し（ステップ30

1)、該キーワードの部分画像特徴量表現を利用して画像データベースの部分画像との類似検索を実行する(ステップ302)。検索結果の部分画像は表示装置に表示される。次に、利用者は、この検索結果の部分画像の中から、該キーワードに関係すると思われるもののうちから未だ該キーワードが設定されていないものを指定する(ステップ303)。この結果、画像データベースのデータについて、該指定された部分画像のキーワードとして該キーワードが追加され(ステップ304)、該指定された部分画像を含む親画像のキーワードとして該キーワードが追加される(ステップ305)。この場合も、追加されるキーワードは、最初にキーワード画像特徴量表現データベースの検索に利用したものと同一でなくともよく、利用者が追加すべきキーワードを別に入力することでもよい。

【0041】このように部分画像にキーワードを追加した画像データベースを利用して、再び図15に示すようにキーワードの部分画像特徴量表現を作成し、キーワード画像特徴量表現データベースに蓄積する。

【0042】新規の画像データを画像データベースに登録する場合、図15の下方のようなキーワードの部分画像特徴量表現を蓄積したキーワード画像特徴量表現データベースを利用することにより、次のようにして、図14の右側のような画像/部分画像キーワード付きデータを自動的に構築することができる。まず、新規に登録する画像、その部分画像の特徴量を抽出して、画像の特徴量セット、部分画像の特徴量セットを取得する。次に、この取得した画像の特徴量セットとキーワード画像特徴量表現データベースにおける各キーワードの画像特徴量表現の特徴量セットとの類似度を算出して、類似度の大きいもののキーワードを該画像に付与する。次に、同様に、取得した部分画像の特徴量セットとキーワード画像特徴量表現データベースにおける各キーワードの部分画像特徴量表現の特徴量セットとの類似度を算出して、類似度の大きいもののキーワードを該部分画像に付与する。

【0043】これを実現するシステムは、基本的に図13に示したような画像検索システムに特徴量抽出部を追加することで可能である。また、部分画像にキーワードを付加することで、図13で説明したと同様の処理で、部分画像の検索が可能である。

【0044】以上、本発明の種々の実施形態について説明したが、本発明は、ここで述べたものに限定されるものでないことは言うまでもない。また、ここで述べたような特徴量の重要度算出手順、あるいはキーワード画像特徴量表現データベース作成手順や画像検索手順は、コンピュータで実行可能なプログラム言語で記述し、あらかじめコンピュータで読み取り可能な記録媒体、例えば、フロッピー(登録商標)ディスク、CD-ROM、メモ리카ード等に記録して提供することも可能である。

これをコンピュータにロードすることにより、所期の機能が実現する。

【0045】

【発明の効果】以上説明したように、本発明による主な効果を挙げれば以下の通りである。

(1) 特徴量の重み付けを算出する際に、特徴量空間(特徴量の分布空間)において、次元数で冪開しているため、特徴量空間の次元数の影響を受けにくい。そのため、画像等の検索において、特徴量ごとの重要性に基づいた特徴量の類似性を判定する際に、精度の高い類似性判定を行うことができる。

(2) 画像に適切なキーワードが付与されていないような画像データベースからでも、キーワード画像特徴量表現(特徴量セットと各特徴量の重要度)からなるデータベースを構築できる。さらに、画像を構成する部分画像ごとに適切なキーワードを付与することができる。

(3) キーワードを特徴量に変換して、キーワードと特徴量の両者を用いて画像データベースを検索できるので、適切なキーワードが付与されていないような画像データベースでも精度の高い検索を行うことができる。

【図面の簡単な説明】

【図1】従来の画像検索システムの構成例を示す図である。

【図2】図1の画像データベースのデータ構成を示す図である。

【図3】従来の画像検索システムの別の構成例を示す図である。

【図4】図3の画像データベースのデータ構成を示す図である。

【図5】従来の特徴量の重要度算出方法を説明する図である。

【図6】従来の特徴量の重要度算出の別の方法を説明する図である。

【図7】従来の特徴量の重要度算出の更に別の方法を説明する図である。

【図8】本発明による特徴量の重要度算出方法の原理を示す図である。

【図9】本発明による特徴量の重要度算出手順の一例を示す図である。

【図10】本発明によるキーワード画像特徴量表現データベース作成システムの構成例を示す図である。

【図11】図10の画像データベースのデータ構成を示す図である。

【図12】キーワードの画像特徴量表現作成の概念図である。

【図13】本発明による画像検索システムの構成例を示す図である。

【図14】部分画像へのキーワードの追加を説明する図である。

【図15】キーワードの部分画像特徴量表現作成の概念

図である。

【図16】部分画像にキーワードを追加する方法の一例を示す図である。

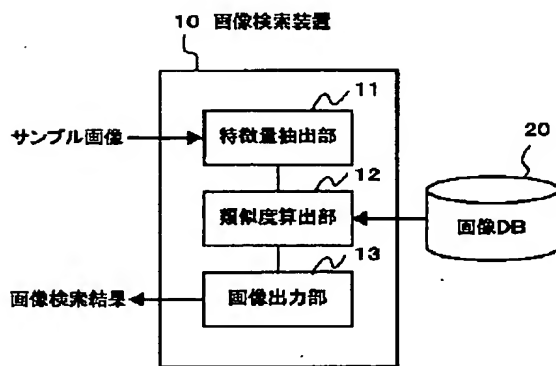
【図17】部分画像にキーワードを追加する方法の別の例を示す図である。

【符号の説明】

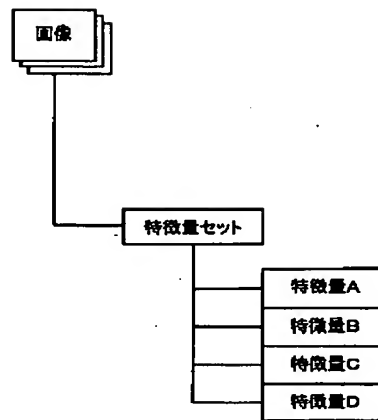
- 110 キーワード画像特徴量表現作成装置
111 キーワード抽出部
112 特徴量セット抽出部
113 特徴量セット算出部

- 114 重要度算出部
115 データ格納部
120 画像データベース
130 キーワード画像特徴量表現データベース
140 画像検索装置
141 キーワード入力部
142 キーワード特徴量変換部
143 特徴量検索部
144 キーワード検索部
145 検索結果統合出力部

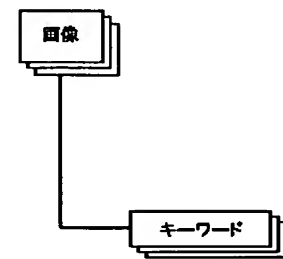
【図1】



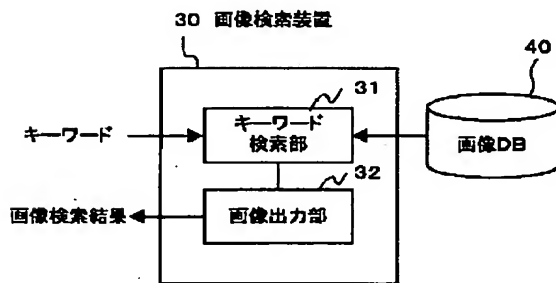
【図2】



【図4】

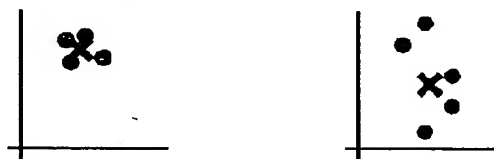


【図3】



【図5】

(A) 分散&標準偏差が小さい (B) 分散&標準偏差が大きい



●：特徴量（ベクトル）

【図6】

【図7】

(A) 外接多次元矩形面積が小さい (B) 外接多次元矩形面積が大きい (A) 外接多次元円面積が小さい (B) 外接多次元円面積が大きい

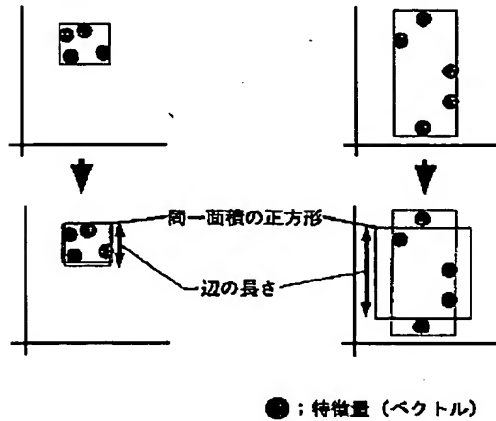


●：特徴量（ベクトル）

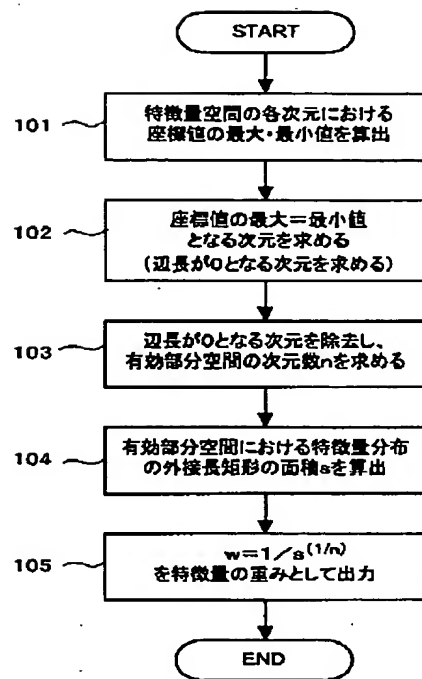
●：特徴量（ベクトル）

【図8】

(A) 外接多次元矩形面積が小さい (B) 外接多次元矩形面積が大きい

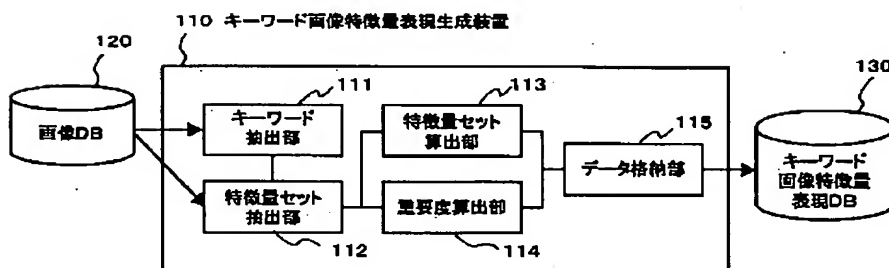


【図9】

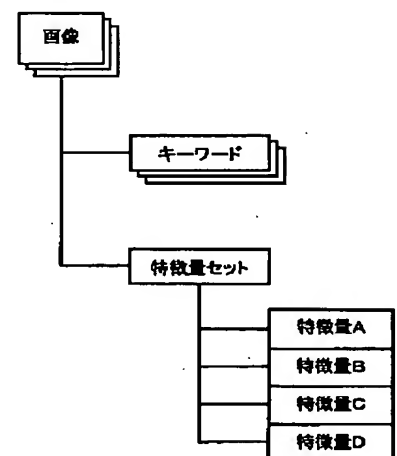


特徴量重要度の算出手順

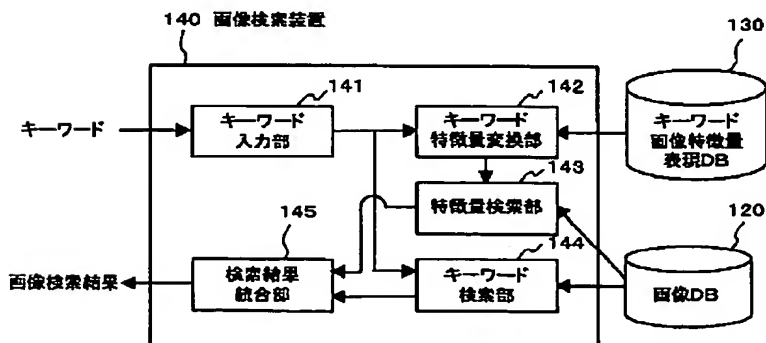
【図10】



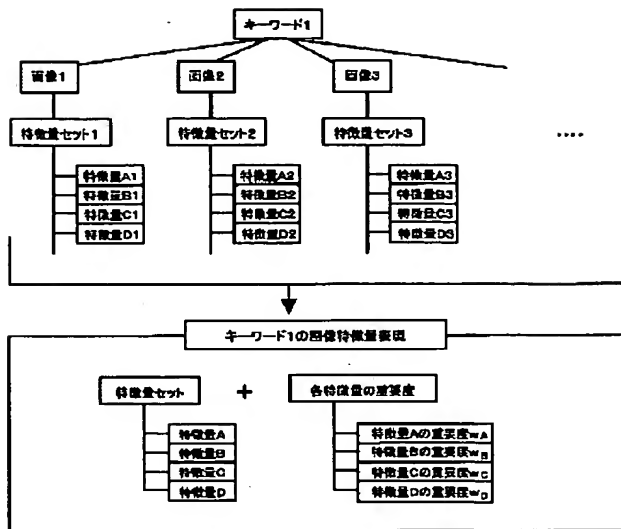
【図11】



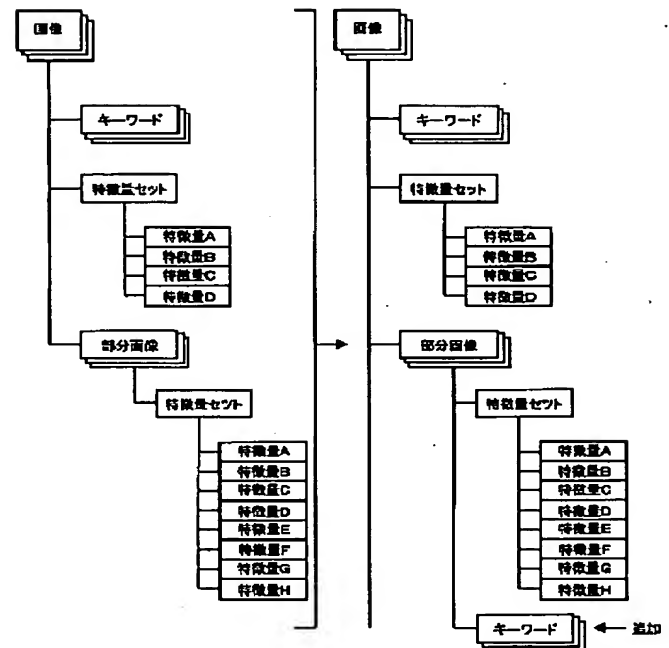
【図13】



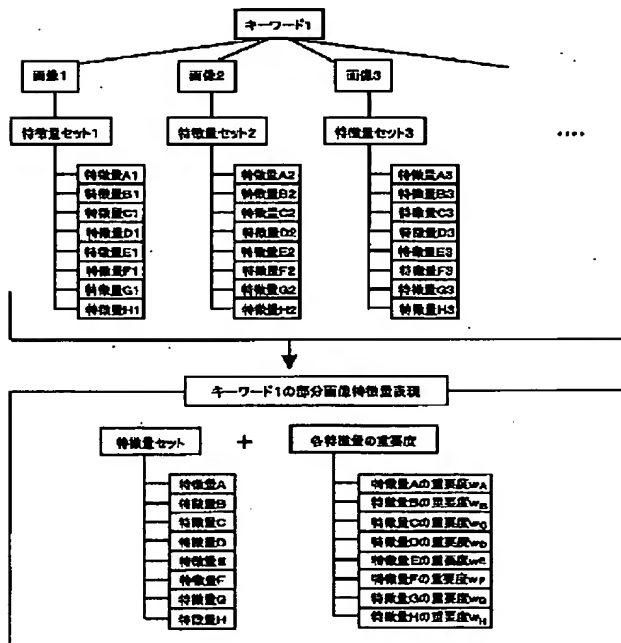
【図12】



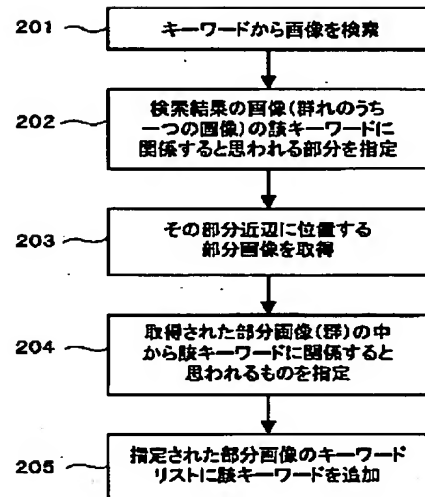
【図14】



【図15】

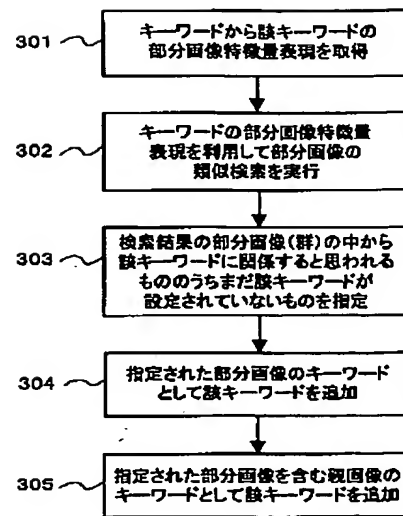


【図16】



部分画像にキーワードを付ける方法(1)

【図17】



部分画像にキーワードを付ける方法(2)